DOCUMENT RESUME

ED 471 757 SE 067 016

AUTHOR	Thompson, Tony; Newsome, Kolouia
TITLE	Is State Assessment a Viable Tool for Reflection of Classroom Assessment?
PUB DATE	2002-00-00
NOTE	11p.; In: Proceedings of the Annual Meeting [of the] North American Chapter of the International Group for the Psychology of Mathematics Education (24th, Athens, GA, October 26-29, 2002). Volumes 1-4; see SE 066 887.
AVAILABLE FROM	ERIC/CSMEE Publications, 1929 Kenny Road, Columbus, OH 43210-1080. Tel: 800-276-0462 (Toll Free).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Academic Achievement; Evaluation; *High Stakes Tests; *Mathematics Education; Secondary Education; *Thinking Skills

ABSTRACT

North Carolina uses Dimensions of Thinking (Marzano et al., 1988) as a framework to incorporate higher order thinking in their End-of-Course exams. The goal of this research was to explore whether this effort could also help facilitate the inclusion of higher order thinking in the classroom. Based on reviewing state EOC-related documents and interviews with state officials, a modified teacher-friendly version of the state's assessment framework was created for classroom use. In Spring 2002, over 60 teachers "recreated" the EOC test item development process and analyzed their current classroom assessment practices using the framework. Most teachers found the framework useful; however, a variety of problems including lack of time, lack of "fit", a restricted context, and low state expectations all worked against using the framework effectively in the classroom. (Author)

IS STATE ASSESSMENT A VIABLE TOOL FOR REFLECTION ON CLASSROOM ASSESSMENT?

Tony Thompson
University of Alabama
athompso@bamaed.ua.edu

Kolouia Newsome
East Carolina University
newsomek@mail.ecu.edu

North Carolina uses *Dimensions of Thinking* (Marzano et al., 1988) as a framework to incorporate higher order thinking in their End-of-Course exams. The goal of this research was to explore whether this effort could also help facilitate the inclusion of higher order thinking in the classroom. Based on reviewing state EOC-related documents and interviews with state officials, a modified teacher-friendly version of the state's assessment framework was created for classroom use. In Spring 2002, over 60 teachers "recreated" the EOC test item development process and analyzed their current classroom assessment practices using the framework. Most teachers found the framework useful; however, a variety of problems including lack of time, lack of "fit," a restricted context, and low state expectations all worked against using the framework effectively in the classroom.

Introduction

State testing has grown significantly in the US over the past decade. Many critics argue that external, state-mandated testing is detrimental to the quality of instruction and assessment in the classroom by (a) narrowing the curriculum, (b) taking flexibility away from teachers to meet the needs of their students, and (c) forcing teachers to "teach to the test." (Kohn, 2000; McNeil, 2000) However, others propose that if state tests focused more on higher order thinking skills, then these tests might actually help teachers improve classroom instruction and assessment by encouraging teachers to include these thinking skills in the classroom. (Yeh, 2001)

North Carolina (NC) is one of many states with yearly testing, and a goal of promoting higher-order thinking skills. The North Carolina Department of Public Instruction (DPI) officially uses *Principle and Standards for School Mathematics* (NCTM, 2000) and *Dimensions of Thinking* (Marzano et al., 1988) as a theoretical framework to develop test items for their End-of-Course (EOC) exams (DPI, 1999; Munk, 2001). The goal of this research was to explore whether a state's effort to assess higher order thinking skills could help facilitate the inclusion of higher order thinking in the classroom. And in particular, to ascertain if NC's assessment framework was a viable tool for improving higher order thinking in classroom assessment.

NC's Use of *Dimensions of Thinking*

Dimensions of Thinking includes: (1) metacognition; (2) critical and creative thinking; (3) thinking processes; (4) core thinking skills; and (5) the relationship of

content-area knowledge to thinking. Although NC DPI encourages teachers to incorporate all five dimensions in their teaching, only the "core thinking skills" are used as basic organizers in the development of EOC test items. The eight core thinking skills listed in *Dimensions of Thinking* (*DOT*) include: (1) focusing; (2) information-gathering; (3) remembering; (4) organizing; (5) analyzing; (6) generating; (7) integrating; and (8) evaluating.

The seven thinking skills used by NC DPI to construct EOC test items include knowledge, organizing, applying, analyzing, generating, integrating, and evaluating. In NC DPI's framework, *focusing*, *information-gathering*, and *remembering* from *DOT* are combined into a single thinking skill called "knowledge." NC DPI also added the additional thinking skill of *applying* not found in *DOT*. Knowledge and applying appears to originate from Bloom's Taxonomy which includes knowledge, comprehension, application, analysis, synthesis, and evaluation. Despite combining the thinking skills of *focusing*, *information-gathering*, and *remembering* into the category "knowledge," *remembering* is the only aspect of this thinking skill used in labeling test items as *knowledge* in NC's EOC exams. In this case, the actual use of this term is more consistent with Bloom's Taxonomy than *DOT*.

Methodology

To frame the professional development experiences for this project, we utilized the ideas of Bryant and Driscoll (1998) to design and facilitate workshops as teachers explored the potential relationship between state and classroom assessment. The "guiding principles for design" included: (a) *purposefulness*: this project was defined around a specific goal / objective with an assessment activity framework specifically created for this project, (b) *connection to the concerns of participants*: this project grew out of our discussions with teachers and their desire to learn more about state testing; (c) *opportunities for exercising judgment*: the project incorporated opportunities for teachers to exercise, share, and discuss their professional judgment through the development and evaluation of state test items; and (d) *opportunities for reflection*: the project was designed to encourage participants to reflect individually and in groups during and after the workshop and relate what they've learned to the classroom.

In Fall 2001, the authors met with DPI officials to discuss the process and theoretical framework (*Dimensions of Thinking*) used for constructing mathematics test items for EOC exams. Shortly thereafter, several teachers agreed to recreate the test item development process in a trial workshop. Using feedback from this process and sharing this result with NC DPI officials, the state's framework was modified to make it more teacher-friendly by grouping the seven thinking skills into three broad categories. In January and February 2002, over 60 teachers participated in two different workshops which involved both analyzing and writing EOC test items. In March 2002, the state assessment framework was modified once again by adding a narrative section to the category descriptors (a more detailed description of the evolution of the assessment framework is provided below).

In creating the workshops the main objectives were to: (a) share the state's process and theoretical framework for creating EOC test items; (b) engage teachers in re-creating the test item development process themselves; and (c) help teachers analyze and compare their current classroom assessment practices with the EOC exams. Following an overview of the process used by DPI to develop test items, each workshop had three small-group activities each followed by whole group discussions.

Activity I: Teachers were given "released" state test items in order to categorize them by thinking skill and level of difficulty. They compared this to the state's categorization for these test items.

Activity II: Teachers wrote their own test items covering a wide range of thinking skills and levels of difficulty. Each group presented their test items for whole group discussion and feedback.

Activity III: Teachers brought copies of their own classroom tests to the workshop. They analyzed their tests by categorizing the questions by thinking skill and level of difficulty and compared this to the proportion of questions found on the EOC exams.

Of these sixty teachers, twenty agreed to continue participating in additional follow-up discussions as they used the assessment framework in their classrooms. For the remainder of the spring semester, they submitted copies of their exams as well as reflections on their experiences in using the assessment framework.

Modifying the State's Assessment Framework for Use by Teachers

Teachers who participate in the development and review of EOC test items describe the process as a rewarding professional development experience. The purpose of this research project was to provide opportunities for teachers to recreate the test item development process to see if it helped them incorporate these higher order thinking skills into the classroom. Our original effort to use the framework "as is" in a "trial workshop" met with little success. Teachers found it too awkward and impractical to use so many different types of thinking skills. In particular, there were large disagreements among teachers when asked to categorize test items by thinking skill. Despite these disagreements, teacher categorizations often fell along similar lines. When reviewing test items, teachers generally grouped "knowledge" items with simple or "one-step" application" items. Similarly, "organizing" was often combined with more complex or "multi-step" application items. And finally, "analyzing," "generating," "integrating," and "evaluating" (referred to as the higher order thinking skills) were often used to classify the same test item.

As a result, we modified the state's assessment framework by grouping the seven thinking skills into three broad categories: *Category I*: Knowledge – Applying (simple or "one-step"); *Category II*: Organizing - Applying (more complex or "multi-step");

and *Category III*: Analyzing, Generating, Integrating, and Evaluating. After two further workshops with teachers and discussions with DPI officials, each thinking skill category was provided with a narrative based on how NC DPI used the thinking skills when developing EOC test items and how teachers often interpreted the thinking skills categories. These descriptions are based on the familiarity a student has to a question and whether or not a student is likely to have been taught an algorithm or procedure to solve the problem. The narrative for each of the categories is as follows:

Category I: Knowledge – Applying ("one-step"). These are recall questions and simple applications. The problems are very familiar to the student and the student should already know an algorithm or procedure to solve the problem.

Category II: Organizing - Applying ("multi-step"). These are more complicated questions and applications, but the student should already know an algorithm or procedure to solve them. The student may also be asked to apply a known procedure to a new situation or the student may need to first organize the information in order to apply a known procedure.

Category III:Analyzing, Generating, Integrating, & Evaluating (Higher Order Thinking Skills). How to solve these questions should *not* be immediately apparent. Students will have to think through the problem before choosing a procedure or will have to create their own procedure / method to solve the problem.

The following are examples of released test items and accompanying thinking skills from the NC DPI website.

Category I

- Simplify: $(6x^3y^3 - 12x^3y^2 + 8x^6y^5) \div 2xy$
- Solve: $2x + 1 = 5$
- Which of the following is a characteristic of a chord?

Category II

- Express, in terms of x, the mean of $(4x^2 - 6)$, $(2x^2 - 3x)$, $(^-13x + 3)$
- The US balance of trade is calculated by subtracting imports from exports. If $0.71x^2 + 2.15x + 67.53$ models exports and $0.82x^2 + 6.42x + 55.07$ models imports for the period 1970 – 1998 ($x = 0$ for 1970), find the algebraic expression for the balance of trade.

Category III

- A soup company decides to increase the height of its cans by 30% but to keep the volume the same. Approximately how much must the radius of the can be decreased to keep the volume constant?

- For the line $y = 3x + 5$: If the y-intercept moves to 7 and the slop remains unchanged, how does the x-intercept change?

In sharing this categorization scheme with DPI officials, they stated that 40% of EOC test items fall in Category I; 20% in Category II, and 20% in Category III. Level of difficulty is another aspect of test item analysis conducted by NC DPI. The level of difficulty of a test item is identified quantitatively through field-testing and qualitatively by teachers. In general, DPI defines easy questions as those 70% of the students would answer correctly; medium questions are those 50% to 60% of the students would answer correctly; and hard test questions are those 20 or 30% of the students would answer correctly. On NC EOC exams, approximately 25% of the items are written at the easy level; 50% at the medium level; and 25% at the hard level.

Thinking skill and level of difficulty are clearly related, but not synonymous. If students rarely have the opportunity to experience Category 3 questions, then all of these questions will be difficult for students to answer. To help teachers differentiate between the thinking skill and the level of difficulty of a test item, an additional framework was created. This framework (in table form) included all possible combinations of thinking skill vs. level of difficulty including the state's percentage goals for each of the nine possible thinking skill–level of difficulty problem types. Teachers used this framework to classify test items from released EOC items or for items they wrote themselves and to place the items into one of the nine categories.

Teachers Use of the Framework in the Classroom

Twenty teachers agreed to use the framework in their classrooms by revising a previous exam and to write a new exam for a future test. Below are examples of Category III questions teachers included on exams that they previously had not. Some teachers took existing questions and extended / modified them to a Category III question. For example,

(Original): "Find the x-intercepts of the function $f(x) = x^3 + 3x^2 + 2x$."

(Revised): "Explain the connections between the x-intercepts of the function $f(x) = x^3 + 3x^2 + 2x$ and the solution to the equation $x^3 + 3x^2 + 2x = 0$."

(Original): "Angela walked along the diagonal of a 30m by 16m rectangular field. How far did she walk?

(Revised): "Angela took a shortcut by walking along the diagonal of a 30m by 16m rectangular field. How much farther would she have had to walk if she had walked along the edges of the field?"

Another common strategy teachers used to write Category 3 questions involved students explaining their answers. For example:

- "Use the discriminant to predict the number and type of solutions of the equa-

tion: $9x^2 + 6x = -1$. Use a graphing calculator to check your solution and explain/describe how the graph verifies your answer."

- "In which of the following triangles can the Pythagorean Theorem be used to find the missing length of a side? Tell why or why not."
- "Can a right triangle be isosceles? Explain."
- "Solve $x^2 + 6x + 2 = 0$. What solution method did you use? Was this the only choice that seems reasonable? Explain."
- "Explain how natural logarithms are different from and similar to common logarithms."
- "Explain why a vertical line has no slope."

Another strategy used by teachers to create Category III questions was reversing problems. That is, give the students the answer and ask them to generate possible questions. For example:

- "Create a word problem that could be solved using the equation $x - 6 = 14$."
- "Find two different equations with roots of –2, 3."

Additional strategies used by teachers to create Category III questions included asking students to

- find an error in a worked problem, then explaining and correcting the error.
- classify equations by type, or comparing/contrasting equations.
- solve problems in more than one way then comparing/contrasting these solution techniques.

Difficulties in Teachers Using the Framework

Most teachers found our modified version of the state's assessment framework a useful tool to reflect on tasks and thinking skills used to solve these tasks. In interviews with teachers, most explained that they previously only used two criteria when creating test items: content coverage and degree of difficulty. However, this framework provided them with a new way to analyze/select/generate assessment items. Nonetheless, there were four problem areas that made it difficult for teachers to use the framework in their classroom: lack of time; lack of "fit;" restricted context; and low state expectations.

Lack of Time

The first area of difficulty was simply a lack of time. Teachers expressed concern over how long it took to write and grade tests which included higher order thinking skills as well as the amount of time it took to prepare students to be able to answer these types of questions on an exam. In the context of the problems discussed below, many of the teachers expressed concern over whether it was time worth spending.

Lack of Fit

The state's identification of test items occasionally did not match the *Dimensions of Thinking* framework nor was the state's own classification scheme consistently applied. This made it difficult for teachers to identify questions by thinking skill. Although many of the questions matched thinking skills from *Dimensions of Thinking*, others did not. Our first two examples[1] are mis-categorized by NC DPI as *evaluation* (category 3) questions:

Example 1: Which expression is equal to $\sqrt{169}$?
a. $\sqrt{121} + \sqrt{9}$ b. $\sqrt{196} - \sqrt{1}$ c. $\sqrt{256} - \sqrt{36}$ d. $\sqrt{144} + \sqrt{16}$

Example 2: Given $y = {}^1/_3x + 2$ and $y = 2x - 13$, which ordered pair below is valid for both equations?
a. $(-2, -1)$ b. $(9, 5)$ c. $(-{}^4/_3, 0)$ d. $(3, 16)$

In these cases, the term "evaluation" as commonly used in mathematics to "find the value of" is incorrectly used as synonymous with "evaluation" as a thinking skill defined in *Dimensions of Thinking* (evaluation: assessing the reasonableness and quality of ideas).

In *Dimensions of Thinking*, integrating refers to connecting and combining information efficiently into a cohesive statement; changing existing knowledge structures to incorporate new information. Nonetheless, the following question was mis-categorized as a Category 3 "integration" question: The length of a barn is 58 feet and the width is 32 feet. Find the area of the barn. (choices not included here).

Sometimes, test item classification was inconsistently used by NC DPI. On one Algebra I EOC exam in 1998, the question "Which of the following is a rational number?" was classified as a "knowledge" (category 1) question while the exact question was listed as an "organizing" (category 2) question on another Algebra I EOC exam. Similarly, some questions identified as "knowledge" (category 1) on one test were categorized as "analyzing" (category 3) on another despite being the exact same question.

This lack of fit often led teachers to believe they were already teaching for higher order thinking even though "routine" textbook problems continued to dominate their exams ... including those they labeled as Category 3. For example, teachers often categorized question such as "Graph $y = 2x - 3$" and "Find an equation with slope 2 and y-intercept 3" as category 3 (Generating) questions. Although these problems have routine solutions, it was actually consistent with NC DPI's use of the term "generating." An example EOC test item classified as generating included: The sum of a number and two is five. What is the equation for this statement? Answer among four choices was $x + 2 = 5$. However, it is not clearly apparent that this fits with the definition of *generating* in *DOT* where it is described as producing new information, meaning, or ideas; going beyond available information to identify what reasonably

may be true; anticipating next events, or the outcome of a situation; explaining details, examples, or other relevant information.

Restricted Context

NC EOC tests are limited to multiple choice questions only. This makes it very difficult to use *Dimensions of Thinking* as originally intended. For example, it can be difficult to construct higher order thinking questions that require students to *evaluate* or *generate* using a multiple choice format. As a result, in the workshops, we asked teachers not to limit themselves to multiple choice questions and encouraged them to broaden its use for all types of classroom assessment/activities.

In addition, teachers recognized that due to the restricted context of the exam, many of the questions could be solved using different cognitive skills ... depending on the knowledge and experience of the student. Teachers felt there was no way to tell if a student was required to use a certain thinking skill. Problems initially non-routine when learned in class may become routine on a test. Although a student may have used higher-order thinking skills to learn the mathematics, eventually the problem may become "routine" and higher-order thinking skills no longer needed on the EOC exam. For example, the following question could be reduced to "find the vertex" by many students; yet, it was classified as an "integrating" question by NC DPI:

> The height in meters of a ball thrown vertically upward is given by the function $h = 20t - 4t^2$. At what time will the ball reach its maximum height and what will be the height at that time?
>
> a. 4 seconds; 25 meters b. 4 seconds; 16 meters
>
> c. 3 seconds; 45 meters d. 3 seconds; 40 meters.

In addition, in the problem above, students are given possible answers. Therefore, students can also substitute the time values to find the corresponding heights. Although this problem does require knowledge of functions, its potential to elicit "higher order" thinking is limited.

Low Expectations

The vast majority of teachers in our workshops stated that they were unaware of the "thinking skills" framework used to construct EOC items (in particular, none of the teachers had heard of *Dimensions of Thinking*). However, most teachers felt their students were doing fine on the Algebra I EOC exam despite their lack of knowledge of the higher order thinking skills. In discussions with DPI officials regarding the Algebra I EOC exam, students often are not required to answer correctly more than 50% of the questions to pass -- implying a student can miss all "category three" questions and still pass the exam. Thus, the low score needed to pass the exam does not "require" students to learn category 3 questions nor do teachers have to prepare their students to answer these types of questions. (Relatedly, as discussed above, many

category 3 questions were mislabeled as "higher – order" thinking questions further lowering the higher order thinking required on the Algebra I EOC exam). [Of note, deeper investigation into the EOC Geometry exam shows higher expectations and a better fit / consistency with *Dimensions of Thinking*. Not surprising, Geometry EOC scores in NC are lower than Algebra I EOC scores.]

Teachers Comments on the Usefulness of the State's Assessment Framework

Despite the diverse interpretations and uses of the NC DPI assessment framework, the general consensus among teachers was that they felt it was useful to have a better understanding of how EOC test questions are created and that they felt that they would make at least a few modifications to how they taught and test so their students would be better prepared for the EOC exam. Several teachers commented that after using the state's assessment framework, they realized they needed to make changes not only to classroom assessment, but instruction as well. As one teacher commented, "When I categorized my test using the framework, I found that 95% of my questions were knowledge [Category I] questions. My new (revised) test is much better because it has a variety of thinking skill levels." She later continued, "To help my students become successful on this test, I would need to implement more high level thinking as part of my daily instruction so that my students could gain more experience in solving problems of this type."

Another teacher, after commenting that her students did very poorly on the "revised" test, explained, "But I did gain some valuable insights from this test. I realized that my style of teaching really does sell short my students' abilities as problem solvers. I try to use discovery type activities, but that is where I seem to stop. Everything after that becomes traditional and rote, especially in my assessments." In general, although the framework was used and interpreted in diverse manners, teachers stated that the framework was a useful starting point to help them select and create tasks using "thinking skills" as a criterion.

Summary

The goal of this research was to explore whether a state's effort to assess higher order thinking skills could help facilitate the inclusion of higher order thinking in the classroom. And in particular, to ascertain if NC's assessment framework was a viable tool for improving higher order thinking in classroom assessment. Our goal was to develop a framework that teachers would find practical in analyzing their classroom assessment practices by helping them introduce "cognitive goals" into their teaching. This project shows that a state's assessment framework can be a viable tool for reflection on classroom assessment. However, the impact on teachers was diverse, and its long-term impact is unclear. There is little evidence to suggest that the state's effort to encourage higher order thinking skills on their exams has had much effect in the classroom without additional training on using the state's assessment framework. Addi-

tional research is needed to continue to follow these teachers to ascertain the impact of their knowledge and use of the state's assessment framework on their teaching and eventually, on the success of their students.

Note

[1]Some of the examples used in this section are similar (but *not* the same) to those found on the NC Algebra I 1998 EOC exam. The 1998 exam questions are secured by the state and used for EOC practice at the district level.

References

Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools*. Portsmouth, NH: Heinemann.

Marzano, R. J., Brandt, R. S., Hughs, C. S., Jones, B. F., Presseisen, B. Z., Stuart, C., & Suhor, C. (1988). *Dimensions of thinking.* Alexandria, VA: Association for Supervision and Curriculum Development.

McNeil, L. M. (2000). *Contradiction of school reform: Educational costs of standardized testing.* New York: Routledge.

Munk, T. (2001). *North Carolina thinking skills: An introduction.* Beacon, NC: Learn.

Department of Public Instruction, Division of Accountability Services, Public Schools of North Carolina, State Board of Education, (1999). *Understanding North Carolina tests: Thinking skill level (Assessment Brief, Vol. 6, No. 4.).* Raleigh, NC: Author.

NCTM (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Yeh, S. (2001). Test worth teaching to: Constructing state-mandated tests that emphasize critical thinking. *Educational Researcher*, *30* (9), 12-17.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

X	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").